

Использование параданных в социологических исследованиях

Аннотация

В статье рассматриваются параданные как особый тип данных, собранных во время опроса о самом процессе сбора информации, таких как поведение респондентов, характеристики условий их проживания, реакции на вопросы, время ответа и др. Приводятся примеры использования параданных в социологических исследованиях.

Ключевые слова: *параданные, метаданные, неответы*

В течение последнего десятилетия все больше опросов проводят с использованием компьютерной техники, причем не только для анализа результатов опросов, но и для сбора информации, то есть получают распространение компьютерно опосредованные методы сбора и анализа информации. В итоге возникает возможность фиксировать, накапливать и анализировать не только сами ответы респондента, но и другие данные и результаты тех или иных процессов, сопровождающих собственно опрос, в частности сведения о времени на размышления над ответом, эмоциональной окраске, внешнем виде респондента, причине отказа от анкетирования и т.п.

Вследствие этого появляются новые возможности для использования метаданных, то есть данных о данных исследования — от формулировки вопросов и альтернатив, времени, места проведения опроса до информации об удобном времени для интервьюирования, отзывах соседей о респонденте и пр.

Итак, важность использования метаданных очевидна. К сожалению, в Украине подобную информацию пока редко используют в социальных исследованиях.

Сформулирую в общих чертах, для чего могут понадобиться метаданные. В социальных науках такие данные применяются по нескольким причинам [Gregory, 2009: p. 7]:

- для уверенности в том, что информация в достаточном объеме доступна для пользователей, для правильного понимания и использования данных. Без подробной документации и описания исследователь не сможет адекватно интерпретировать и понять суть данных;
- для облегчения поиска и доступа к данным по потребительским признакам: ключевым словам, теме и цели исследования, описанию выборки и генеральной совокупности и др. Даже самые лучшие данные в мире окажутся бесполезными, если о них никто не знает и не использует их;
- для поддержания длительного хранения данных с уверенностью в том, что важная информация остается при них для дальнейшего использования или для перекодирования в другие форматы;
- для обеспечения и поддержания обмена информацией между индивидами или институтами.

Чем более детальной и полной является сопроводительная документация об исследовании, его методологии, методах, генеральной совокупности, выборке, тем удобнее использовать его результаты исследователям.

Одной из разновидностей метаданных являются параданные — информация о самом процессе опроса.

Цель данной статьи — охарактеризовать концепт параданных как особого вида метаданных и привести примеры их использования в социологических исследованиях на основе анализа международного опыта.

Социальные исследования во многих странах имеют хорошо развитую практику использования метаданных, поскольку сами данные исследования — это ответы респондентов, собираемые, обрабатываемые, анализируемые и используемые в ходе исследования, а метаданные — это вся документация о данных. Современные немецкие исследователи А.Грегори и его соавторы [Gregory, 2009: p. 2] вводят следующую типологию метаданных:

Структурные метаданные описывают структуру данных: имена переменных; какие значения переменных появляются в каких колонках; какая строка представляет какой случай; имеются ли иерархические связи и т.п. Эту информацию должен содержать паспорт анкеты (статистический пакет ОСА) или вкладыш “Переменные” (статистический пакет SPSS).

Описательные метаданные представляют собой информацию о методологии, выборке, качестве измерения. Такая информация должна быть в программе исследования.

Административные метаданные — это информация, создающаяся самим процессом управления данными в ходе их сбора, обработки, публикации и т.д. К таким данным относятся данные о правах доступа к результатам, юридические документы исследования, информация о формате файла хранения данных и месте его расположения и др.

Поведенческие метаданные (которые называются *параданными*) являются информацией о реакции и поведении исследователей при использовании данных исследования, а также поведении интервьюеров и респондентов в ходе опроса. К параданным относятся оценки интервьюером внешнего вида респондента, реакция респондента на вопрос, состояние домохозяйства, впечатления соседей, причины отказов и т.п.

В дальнейшем мы детальнее остановимся только на четвертом типе, который называется параданными, чему и посвящена статья.

Термин “paradata (параданные)” впервые применил Мик Купер, американский социолог в 1998 году в Далласе на семинаре JSMASA (Joint Statistical Meetings of the American Statistical Association) при обсуждении доклада [Couper, 1998] “Измерение качества опросника в среде компьютерно опосредованного сбора данных” и на 53-й конференции AAPOR (American Association for Public Opinion Research) [Hansen, 1998]. Внесенное Купером предложение об использовании данных о процессе опроса, автоматически сгенерированных компьютером во время CAPI (Computer Assisted Personal Interviewing), позже было расширено для других аспектов самого выборочного процесса и прочих методов сбора данных — CATI (Computer Assisted Telephone Interviewing) и PAPI (Paper Assisted Personal Interviewing).

Использование данных о процессе сбора информации позволяет улучшить качество результатов и уменьшить погрешности исследования, проанализировать причины неответов, оценить их влияние на результаты и определить оптимальные характеристики для дальнейших исследований, таких как время проведения исследования, отбор респондентов и т.п. [O’Reilly, 2009; Kreuter, 2010(a); (b)], что отмечалось на некоторых международных конференциях (International Workshop on Household Survey Non-response, European Survey Research Association (ESRA), American Association for Public Opinion Research (AAPOR) и др.).

Стандартного однозначного определения понятия “параданные” пока не существует. Сначала параданные определялись как данные, автоматически записанные компьютером в ходе проведения интервью [Hansen, 1998], но со временем значение этого термина было расширено и стало охватывать данные, записанные интервьюером в процессе проведения опроса либо полученные из других источников, например аудиозапись телефонного интервью [Couper, 1998].

В таблице 1 приведены разновидности происхождения параданных в зависимости от типа опроса.

Таблица 1

Источники получения параданных

Тип исследования	Местонахождение параданных	Разновидность записи
CAPI	Сервер, персональный компьютер	Автоматически компьютером
CATI	Сервер	Автоматически компьютером, частично интервьюером
PAPI	Контактный лист	Интервьюером

Важно не столько дать четкую интерпретацию понятия “параданные”, сколько понять сам концепт полученных во время опроса данных о его ходе.

Одним из типов данных, которые считаются параданными, являются записи звонков, собранные в процессе опроса выборки путем телефонного интервью. Время звонка, дата, как и результат звонка (отказ, интервью, не поднимают трубку, недействительное, не законченное и др.), являются доступными данными, автоматически записанными системой CATI при попытке провести интервью. По этим данным можно, например, скорректировать

время проведения интервью с целью уменьшения количества отказов или ситуаций отсутствия абонента [Laflamme, 2008]. В США, к примеру, проводят исследование (US National Survey of Family Growth), где параданные используются для ежедневной централизованной коррекции схемы опроса по телефону [Lepkowski, 2009]. Однако в большинстве исследований параданные анализируются после сбора информации.

Второй набор данных, относящихся к параданным, также собирают на начальной фазе установления контакта и привлечения к опросу типичных респондентов. Отличие рассмотренных выше параданных заключается в том, что здесь сам интервьюер собирает эти данные, которые должны охватывать все случаи — как те, когда респондент соглашается участвовать в интервью, так и случаи отказа. Интервьюеры должны фиксировать наблюдения об окружении респондентов, жилищных характеристиках, указывать причины отказа участвовать в исследованиях. Для повышения доли ответов разработаны с учетом этих особенностей целые схемы и системы рекомендаций интервьюерам, например PEDAКSI (Pre-Emptive Doorstep Administration of Key Survey Items) [Lynn, 2003], IAS (Interviewer Attitudes Survey) [Campanelli, 1997] и др.

Примером сбора параданных может послужить исследование National Survey of Family Growth, которое проводится в США. Это многоразовое исследование домохозяйств, в которых для опроса отбирают респондента в возрасте от 15 до 44 лет, с которым проводят интервью в течение 60–80 минут с использованием компьютера. Исследование сфокусировано на аспектах рождаемости и сексуального опыта. Интервьюерам приходилось наблюдать за домохозяйствами и оценивать наличие сексуального партнера противоположного пола у респондента, наличие детей в домохозяйстве. Вместе с тем в анкете также фигурировали вопросы на другие темы. Таким образом, параданные использовались для взвешивания результатов опроса [Groves, 2007].

В Европейском социальном исследовании (ESS) во всех волнах используется контактная форма, предполагающая заполнение интервьюером [Социальное исследование ESS (a), 2008] и призванная обеспечить достаточно полную картину касательно самого респондента, его местожительства, готовности принять участие в опросе и др. Эти данные относятся к параданным и используются как для корректировки самого процесса сбора информации, так и для коррекции влияния неответов. В документе “Response-based Quality Assessment” [Социальное исследование (b), 2009] описана роль параданных, полученных из контактных листов интервьюеров. Сначала собирают сведения обо всех типичных единицах, включая тех, кто ответил на анкету, и тех, кто отказался от анкетирования (такая контактная форма [Социальное исследование ESS (a), 2008] также включает информацию о визуальных наблюдениях интервьюера, его оценку среды пребывания). Это позволяет описывать людей, отказывающихся от опроса, и тех, кто готов его пройти. В случае отказа интервьюер должен оценить готовность респондента участвовать (либо снова отказаться) в следующих или повторных опросах. Именно информация о “неответе” и дает возможность в дальнейшем корректировать модель выборки. После этого обо всех респондентах собираются оценки их жилища (домохозяйства) в плане условий проживания. Кроме того, эти данные собираются в течение всего исследования, поэтому

можно проследить также определенную динамику. Контактные формы постоянно модифицируются и улучшаются ради большего удобства использования их интервьюером.

Эти примеры иллюстрируют возможность как текущих корректировок программы исследования для уменьшения неответов, так и модификации программы последующих исследований. Условия появления неответов можно отследить как по респондентам, согласившимся участвовать в исследовании, так и по тем, кто отказался, поскольку параданные фиксируются и для тех, и для других. По ним можно моделировать процесс опроса таким образом, чтобы уменьшить появление неответов.

В автоматизированных опросах есть возможность записывать весь разговор, включая интонацию, мимику, эмоции, а не только фиксировать контактную форму. Эти данные отнесем к третьему типу параданных. Исследования подтверждают, что реакция респондента на приглашение принять участие в опросе свидетельствует о его уровне потенциального согласия на интервью; кроме того фиксируется, как голосовые характеристики интервьюера влияют на согласие на опрос [Groves, 2007]. Аудиозапись (или видеозапись) и полученные таким образом параданные представляют интерес не только на этапе рекрутирования респондентов для проведения опроса, но и для оценки погрешности измерений для каждого из вопросов, ведь эмоциональное состояние влияет на взвешенность респондента. Даже комментарии респондента к вопросам могут повлиять на выбор его ответов — в какой мере его волнует проблема и насколько эмоционально он относится к вариантам ответа и самому вопросу. Таким образом, из аудиозаписи можно получить следующие параданные: изменение интонации респондента при прослушивании альтернатив, изменения вариантов своих ответов на вопрос, комментарии к вопросам и альтернативам, время на размышление, уточнение вопроса; из видеозаписи — еще мимика, сконцентрированность или рассеянность, движение глаз, жесты и т.п. На первый взгляд эти дополнительные данные не существенны, однако в действительности они содержат множество объективных характеристик (то есть специально не контролируемых ни респондентом, ни интервьюером), о чем свидетельствуют некоторые исследования (например, [Conrad, 2010]). К примеру, по движению глаз можно оценить правдивость ответов респондента, по мимике — уровень возбудимости, влияющий на эмоциональность ответов.

Особо следует сказать о сборе параданных при проведении опросов в среде Интернет, где есть возможность осуществить не только анкетирование, но и аудио- и даже видеопрос, об эффективности и возможностях получения информации в рамках которых уже говорили многие исследователи (см. : [Fuchs, 2007]). Прежде всего автоматически собирается информация о подключении респондента к Интернету (IP-адрес, провайдер, регион, тип подключения) и некоторые параметры компьютера (тип браузера, разрешение экрана, операционная система, некоторые дополнительные характеристики компьютера, свободно транслирующиеся в ходе передачи запросов). Автоматически фиксируется время начала и окончания опроса, кроме того, большинство систем для проведения опросов он-лайн имеют возможность фиксировать не только время, которое респондент потратил на каждый из ответов, но даже движение мышкой по экрану в процессе размышлений для определения сосредоточенности на ответе и активности окна с опросником.

При проведении исследований в web (электронная анкета, расположенная на web-странице) параданные можно получать как из серверной части опросника (server side), так и с клиентской (client side). Из сервера получают информацию о том, откуда зашел респондент в опросник (IP-адрес), когда зашел, в котором часу начал опрос, когда его закончил, ответил на все вопросы за один раз или прерывал процесс заполнения анкеты, персональный пароль доступа, адрес, с которого был осуществлен переход на анкету, тип операционной системы, тип браузера. По двум последним параметрам можно оценить быстродействие компьютера и сделать поправку на скорость загрузки для поправки оценки времени, потраченного на ответ. Эти параданные сохраняются на сервере либо в виде таблицы базы данных, либо в виде текстового файла с протоколом. Значительно более интересные параданные можно получить из клиентской части опросника. Клиентская часть опросника является программой, встроенной в само тело анкеты. Эта программа, как правило, написана на Java, php или на другом языке программирования в web и должна собирать дополнительную информацию о ходе опроса. Она может работать только тогда, когда это разрешено компьютером клиента и не является вирусом, шпионской системой или любым другим нежелательным компонентом. С помощью клиентской части опросника можно получать информацию о: времени, потраченном респондентом на ответы на каждый вопрос, о том, изменял ли ответы респондент в ходе опроса, его навигации по вопросам опросника, если это разрешено самим опросником, а также оценить эти параметры в отношении разных типов вопросов и разновидностей подачи альтернатив (кнопки, галочки, выпадающие меню альтернатив, табличные вопросы с ранжированными совместимыми альтернативами и т.п.). Например, оказалось, что выпадающие меню альтернатив (drop-down boxes) менее удобны для респондентов, чем кнопки radio button (набор альтернатив, из которого можно выбрать только один вариант ответа, причем все альтернативы записаны в строку или в столбик) [Sowan, 2010]. Применение этих client side параданных имеет весьма широкий спектр. Бельгийский социолог Д.Хеервег в [Heerwegh, 2004] приводит четыре примера эффективного использования параданных, полученных из клиентской части web-опросника: калибровка индикатора уровня заполненности анкеты как одного из факторов, мешающих исчерпывающему заполнению анкет респондентами; оценка эффекта использования вопросов в разных форматах для повышения удобства фиксации ответов для респондента; тест эффективности различных формулировок вопросов с целью выявления наиболее однозначной и непредвзятой формулировки как самого вопроса, так и альтернатив; оценка компетентности и уровня отношения к проблеме путем оценки времени ответа и изменения ответов (траектории ответов).

Очевидно, что этот перечень примеров далеко не полный, однако он дает направление для изучения возможностей использования параданных в социологических исследованиях.

Итак, мы можем сформулировать несколько главных причин применения параданных:

- текущий анализ процедуры сбора данных и ее коррекция, способствующая экономии средств;
- использование параданных для ремонта выборки после сбора информации, что повышает качество результатов исследования.

Каждая социологическая компания пытается повысить качество собранной социологической информации с минимальными финансовыми затратами. Влиятельным фактором уменьшения репрезентативности выборки является уровень неответов как в анкетировании, так и в телефонном интервью. Национальная канадская статистическая компания Statistics Canada [Statistics Canada, s. a.] выступает одним из инициаторов поиска различных дополнительных возможностей для улучшения качества результатов исследования без увеличения его стоимости. Одним из таких путей является использование параданных при проведении CATI [Lafamme, 2008]. В статистическую базу записываются все события, происходящие в ходе опроса: дата первого и всех последующих звонков, время начала звонка, время окончания звонка, время на каждый ответ, состояние контакта (отказ без согласия на повторный опрос, отказ с переносом времени опроса и др.), ответ на все вопросы в течение одного сеанса или нескольких, место нахождения респондента (в случае звонка на мобильный телефон), эмоциональное состояние респондента и т.п. По этим данным корректируется оптимальное время контакта, манера обращения интервьюера [Conrad, 2010] для установления контакта и т.п. Использование параданных позволило сократить время, потраченное интервьюером на опрос, и количество неответов.

Еще одним применением параданных является попытка обнаружить связь между параданными и ответами и, соответственно, осуществить дальнейшее корректирующее взвешивание данных или неответов [Little, 2005; Maitland, 2008]. В докладе Кретера 18 февраля 2011 года на “Research Agenda for the Future of Social Science Data Collection Workshop” [Kreuter, 2011] приведены корреляционные зависимости между параданными и некоторыми признаками в разных социальных исследованиях. В [Kreuter, 2010(a)] и [Maitland, 2008] также обнаружены зависимости между уровнем отказов и характеристиками респондентов. Например, в [Maitland, 2008] изложены результаты исследования, которое проводилось в США в 2006 году; объем выборки составлял 44264 семьи, из которых были отобраны 31142 (те, в которых был установлен контакт хотя бы с одним из членов семьи). Рассматривались три типа переменных: параданные (z), участие в опросе (p) и переменные интереса (y). К параданным отнесли три типа переменных, которые гипотетически могли бы повлиять на коррекцию неответов: одна группа переменных описывает усилия по установлению контакта с членом семьи (наличие помех для доступа в помещение, отсутствие членов семьи дома и пр.); вторая группа измеряет уровень сотрудничества в домохозяйстве (респондент заявляет, что ему не интересно это исследование, он болен, устал и др.); к третьей группе относится отказ от интервьюирования в течение самого опроса, респондент прерывает интервью, ссылаясь на здоровье или на другие причины).

Сначала анализировались корреляции между переменными групп z , p и y , и $z-p$ корреляции оказались в пределах 0,02–0,41 со средним значением 0,14. Наиболее сильным прогнозом несогласия отвечать на вопросы анкеты была переменная параданных, которая показывала отсутствие заинтересованности и занятость респондента. В то же время корреляция между уровнем контактности и согласием на анкетирование колебалась между 0 и 0,17 со средним 0,07, а наиболее сильным индикатором был факт, когда в домохозяйстве не отвечали на звонок (никого не было дома или изображали отсутствие жильцов). Самым высоким коэффициентом корреляции между переменными z и y был 0,1 со средним значением 0,05.

Затем был применен факторный анализ для переменных, которые отражали сотрудничество и контактность. В таблице 2 приведены результаты связи между факторами, с одной стороны, и участием в опросе и переменными заинтересованности (интереса) — с другой).

Таблица 2

Корреляционная таблица

Наборы переменных	Связь между факторами и участием в опросе	Связь между факторами и переменными заинтересованности	
		диапазон	среднее
<i>Переменные сотрудничества</i>			
Фактор 1: Содержание опросника/приватность	0,28	0,01–0,06	0,03
Фактор 2: Отсутствие времени	0,24	0,01–0,08	0,04
Фактор 3: Враждебность на просьбу пройти анкетирование	0,47	0,00–0,05	0,02
Фактор 4: Проблема доступа в помещение (консьерж, сторож)	0,07	0,00–0,04	0,01
<i>Переменные контактности</i>			
Фактор 1: Проблемы установления контакта	0,18	0,00–0,10	0,05
Фактор 2: Место проживания/проблема отбора	0,02	0,00–0,04	0,02
Фактор 3: Проблемы препятствий (ворота, забор)	0,11	0,00–0,06	0,03

Путем логистической регрессии с переменной “статус участия в опросе” как зависимой переменной и всеми индивидуальными переменными параданных как независимыми переменными была получена оценка склонности к участию в опросе для каждого респондента.

Итак, можно сказать, что использование параданных способно существенно улучшить качество результатов социальных (маркетинговых) исследований без увеличения стоимости самого исследования и сделать возможным более точное моделирование выборочной совокупности. Особенно это касается систем CAPI и CATI, хотя, как, например, в случае ESS, использование PAPI с контактными формами также делает возможным сбор параданных. Использование параданных при проведении web-опросов показывает высокую эффективность для анализа как дизайна анкеты, так и повышения процента ответов.

Таким образом параданные могут быть использованы:

- в ходе пилотажного исследования для моделирования выборочной совокупности и коррекции программы исследования;
- при проведении исследования для программирования дальнейших исследований;
- при проведении исследования (благодаря анализу неответов) для коррекции результатов исследования;
- в процессе анализа данных для перевзвешивания ответов.

Возникает логичный вопрос: в каком формате сохранять параданные (и другие метаданные) так, чтобы он был, с одной стороны, универсальным, а с другой — понятным исследователям? Существует множество стандартов для хранения метаданных для разных отраслей науки и техники [Определение “метаданные”, s.a.]. В социальных науках используются главным образом несколько стандартных моделей метаданных: Data Documentation Initiative (DDI) [Data Documentation Initiative, s.a.], ISO-TS 17369 Statistical Data and Metadata Exchange3 (SDMX) [Statistical Data and Metadata, s.a.], Dublin Core Metadata Initiative (DCMI) [The Dublin Core Metadata, s.a.], ISO/IEC 11179 [Metadata Standards, s.a.] и Neuchâtel models for variables and classifications [Neuchâtel Model, s.a.]. Не буду сравнивать сейчас эти модели, подчеркну лишь, что все они призваны как можно лучше соответствовать хранению метаданных.

К сожалению, остаются неразрешенными этические вопросы использования параданных, ведь в них содержится достаточно субъективная и частная информация, которая не может быть доступной для всех. Следовательно, нужно фильтровать объем и тип параданных для общего доступа так, чтобы не навредить респондентам и не нарушить их прав, то есть путем удаления персональных идентификационных данных.

Разумеется, параданные не являются универсальным механизмом для снижения стоимости исследования и повышения его качества. Это лишь еще одна из возможностей уменьшить влияние неответов на результаты исследования, получить дополнительную информацию для взвешивания самих ответов, оценить модель выборки и пути ее реализации. Далеко не всегда в исследованиях такие данные выделяют в отдельную категорию и называют “параданными”, но тенденция к использованию дополнительной информации для улучшения результатов сохраняется.

Источники

Определение “метаданные” в Википедии [Электронный ресурс]. — Режим доступа: <http://ru.wikipedia.org/wiki/Метаданные>.

Соціальне дослідження ESS, контактна форма 4 раунду 2008 рік [Электронный ресурс]. — Режим доступа : http://ess.nsd.uib.no/streamer/?module=main&year=2009&country=UA&download=%5CFieldwork+documents%5CUA%5C2009%5C04%23ESS4+-+Contact+forms%5CLanguages%5CRussian%5CESS4ContactformUA_rus.pdf.

Соціальне дослідження ESS, Response Based Quality Assessment [Электронный ресурс]. — Режим доступа : http://ess.nsd.uib.no/streamer/?module=main&year=2009&country=null&download=%5CSurvey+documentation%5C2009%5C09%23ESS4+-+Response+Based+Quality+Assessment%5CLanguages%5CEnglish%5CResponse-based+Quality+Assessment+ESS4CF_e02.pdf.

Campanelli P. Can you hear me knocking: An investigation into the impact of interviewers on survey response rates / P. Campanelli, P. Sturgis, S. Purdon // Technical report (Final Report for UK ESRC grant R000235776). — London : The Survey Methods Centre at SCPR (Social and Community Planning Research), 1997.

Conrad F. To agree or not to agree: Effects of spoken language on survey participation decisions / [F. Conrad, J.G. Broome, J. Benki, R. Groves, F. Kreuter, D. Vannette] // 63th Annual Conference of the American Association of Public Opinion and Research (AAPOR). — Chicago (ILL), 2010. — May 16.

Couper M.P. A Measuring Survey Quality in a CASIC Environment [Electronic resource] / Couper M.P. : Invited paper presented at the Joint Statistical Meetings of the American Statistical Association // Proceedings of the Survey Research Methods Section, ASA (1998), Achieving Quality in Surveys, Dallas, August. — P. 41–49. — Mode of access : <http://www.amstat.org/sections/srms/Proceedings/y1998f.html>.

Data Documentation Initiative (DDI) [Electronic resource]. — Mode of access : <http://www.ddialliance.org/>.

The Dublin Core Metadata Initiative [Electronic resource]. — Mode of access : <http://dublincore.org/>.

Fuchs M. Video Web Survey — Results of an experimental comparison with a text-based Web survey [Electronic resource] / M. Fuchs, F. Funke // Proceedings of the Association for Survey Computing's Fifth International Conference on the Impact of Technology on the Survey Process.— Southampton (Eng.). — 2007. — September. — P. 63–80. — Mode of access : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.960&rep=rep1&type=pdf>.

Gregory A. Metadata [Electronic resource] / Gregory Arofan, Heus Pascal, Ryssevik Jostein // RatSWD, working paper № 57, march 2009 P22. — Mode of access : http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_57.pdf.

Groves R. Use of interviewer judgments about attributes of selected respondents in post-survey adjustment for unit nonresponse: An illustration with the National Survey of Family Growth [Electronic resource] / R. Groves, J. Wagner, E. Peytcheva // Proceedings of the Section on Survey Research Methods of the American Statistical Association. — 2007. — P. 3428–3431. — Mode of access : <http://www.amstat.org/sections/srms/Proceedings/y2007/Files/JSM2007-000782.pdf>.

Hansen S.E. Usability Evaluation of the NHIS CAPI Instrument [Electronic resource] / S.E. Hansen, Couper M.P., Fuchs M. // Papers Presented at the 53rd Annual Conference of the American Association for Public Opinion Research. — 1998. — May 14–17, St. Louis, Missouri. — P. 928–933. — Mode of access : http://www.amstat.org/sections/srms/Proceedings/papers/1998_160.pdf.

Heerwegh D. Uses of client side paradata in Web surveys [Electronic resource] / D. Heerwegh // International Symposium in Honour of Paul Lazarsfeld. — Brussels : Belgium, 2004. — Mode of access : <http://www.websm.org/uploadi/editor/Heerwegh20Uses20of20Client20Side20Paradata20in20Web20Surveys.pdf>.

Kreuter F. (a) Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data [Electronic resource] / F. Kreuter, G. Müller, M. Trappmann // Public Opinion Quarterly. — 2010. — Vol. 74, № 5. — P. 880–906. — Mode of access : <http://poq.oxfordjournals.org/content/74/5/880.full.pdf>.

Kreuter F. (b) Paradatam [Electronic resource] / F. Kreuter, C. Casas-Cordero // RatSWD, working paper № 136, april 2010 P16. — Mode of access : http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_136.pdf.

Kreuter F. Using Paradata for Post Survey Adjustments: State of the Science [Electronic resource] / F. Kreuter // Panel on a Research Agenda for the Future of Social Science Data Collection Workshop. — 2011. — February 17–18. — Washington (DC). — Mode of access : <http://www7.nationalacademies.org/cnstat/Kreuter%20Pres.pdf>.

Laflamme F. Data Collection Research using Paradata at Statistics Canada [Electronic resource] / F. Laflamme // Symposium 2008 : Data Collection : Challenges, Achievements and New Directions. — Component of Statistics Canada Catalogue. — № 11-522-X. — Mode of access : <http://www.statcan.gc.ca/pub/11-522-x/2008000/article/10997-eng.pdf>.

Lepkowski J. Use of paradata to manage a field data collection / [J. Lepkowski, R. Groves, W. Axinn, N. Kirgis, W. Mosher] // Proceedings of the Section 6 on Survey Research Methods of the American Statistical Association. — 2009. — August 2. — Joint Statistical Meetings.

Little R. Does Weighting for Nonresponse Increase the Variance of Survey Means? [Electronic resource] / R. Little, S.L. Vartivarian // Survey Methodology. — 2005. — 31. —

P. 161–168. — Mode of access :

<http://www.mathematica-mpr.com/publications/pdfs/weightingnonresponse.pdf>.

Lynn P. PEDAksi : Methodology for collecting data about survey non-respondents [Electronic resource] / P. Lynn // *Quality and Quantity*. — 2003. — № 37 (3). — P. 239–261. DOI: 10.1023/A:1024449208633. — Mode of access :

<http://www.iser.essex.ac.uk/publications/working-papers/iser/2002-05.pdf>.

Maitland A. An Exploration into the Use of Paradata for Nonresponse Adjustment in a Health Survey [Electronic resource] / A. Maitland, C. Casas. — Cordero, Frauke Kreuter // *Proceedings of the Survey Research Methods Section, American Statistical Association (2008)*. Session 381: Paradata, Data Quality, and the National Health Interview Survey (NHIS). — Mode of access :

<http://www.amstat.org/sections/SRMS/proceedings/y2008/Files/301425.pdf>.

Metadata Standards ISO/IEC JTC1 SC32 WG2 [Electronic resource]. — Mode of access : <http://metadata-standards.org/>.

Neuchâtel Model — Classifications and Variables [Electronic resource]. — Mode of access : <http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=14319930>.

O'Reilly J. Paradata and Blaise: A review of recent applications and research [Electronic resource] / O'Reilly J. : Paper presented at the 12th International Blaise Users Conference (IBUC). — 2009. — Latvia. — Mode of access :

<http://www.blaiseusers.org/2009/papers/7d.pdf>.

Sowan A.K. Paradata: A New Data Source From Web-Administered Measures [Electronic resource] / Sowan A.K., Jenkins L.S. // *CIN: Computers, Informatics, Nursing*. — 2010. — Vol. 28. — № 6 (November/December). — P. 333–342. — Mode of access :

<http://www.nursingcenter.com/pdf.asp?AID=1079600>.

Statistics Canada [Electronic resource]. — Mode of access :

<http://www.statcan.gc.ca/start-debut-eng.html>.

Statistical Data and Metadata exchange [Electronic resource]. — Mode of access : <http://sdmx.org/>.